Moshe Buchinsky
Department of Economics
UCLA

Economics 203C
Spring 2003

Lecture Note 5
Estimation of Regression with Endogenous Regressors

$$y_i = x_i'\beta + \epsilon_i, \quad E[x_i\epsilon_i] \neq 0,$$

in the cross section regression context, or

$$y_t = x_t'\beta + \epsilon_t, \quad E[x_t\epsilon_t] \neq 0,$$

in the time series context.

I. INSTRUMENTAL VARIABLES

Let z_i be an $l \times 1$ vector, with

$$E[z_i\epsilon_i] = 0$$

and

$$E[z_i x_i'] = \Sigma_{zx},$$

an $l \times k$ matrix, with $\text{rank}(\Sigma_{zx}) = k \leq l$, so Σ_{zx} is not invertible.

Note that by the Central Limit Theorem (CLT),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{T} z_i\epsilon_i \xrightarrow{D} N(0, V_0).$$

"WEIGHTED" IV ESTIMATOR:

$$\hat{\beta}_{IV}(\hat{\Pi}) = (\hat{\Pi}'Z'X)^{-1}(\hat{\Pi}'Z'y),$$

where $\hat{\Pi}$ is an $l \times k$ matrix, with $\hat{\Pi} \xrightarrow{p} \Pi_0$.

For example: $\hat{\Pi} = (1/n)(Z'X)'A^{-1}$, where $A \xrightarrow{p} \Phi$, a positive definite matrix.

ASYMPTOTIC DISTRIBUTION:

Naturally, the asymptotic distribution of $\hat{\beta}_{IV}(\hat{\Pi})$ depends on Π_0.

$$\sqrt{n}(\hat{\beta}_{IV} - \beta) \xrightarrow{D} N(0, \Lambda),$$

where

$$\Lambda = (\Pi_0' \Sigma_{zx})^{-1} \Pi_0' V_0 \Pi_0 (\Pi_0' \Sigma_{zx})^{-1}.$$

If $\hat{\Pi} = (1/n)(Z'X)' A^{-1}$, then $\Pi_0 = \Sigma_{zx}' \Phi^{-1}$, and

$$\Lambda = (\Sigma_{zx}' \Phi^{-1} \Sigma_{zx})^{-1} \Sigma_{zx}' \Phi^{-1} V_0 \Phi^{-1} \Sigma_{zx} (\Sigma_{zx}' \Phi^{-1} \Sigma_{zx})^{-1}.$$

II. TWO-STAGE LEAST SQUARES

Use $\hat{X}$ as a predictor of X from the regression of X on Z. That is,

$$\hat{\Pi} = (Z'Z)^{-1} Z'X,$$

and

$$\hat{X} \equiv Z\hat{\Pi} = Z(Z'Z)^{-1} Z'X = H_z X,$$

where $H_z = Z(Z'Z)^{-1} Z'$ and $E[\epsilon \mid Z] = 0$.

Hence,

$$y_i = \hat{x}_i' \beta + \epsilon_i^*,$$

where

$$\epsilon_i^* = \epsilon_i + (x_i - \hat{x}_i)' \beta$$

and by construction,

$$E[\epsilon_i^* \mid \hat{x}_i] = 0.$$

The two-stage least estimator of β is given the by

$$\begin{aligned}
\hat{\beta}_{2SLS} &= (\hat{X}'\hat{X})^{-1} \hat{X}' y \\
&= \left(X'Z(Z'Z)^{-1} Z'X \right)^{-1} X'Z(Z'Z)^{-1} Z'y \\
&= (X' H_z X)^{-1} X' H_z y,
\end{aligned}$$

since H_z is idempotent matrix.

INTERPRETATION:

1. We can write the 2SLS estimator as:

$$\hat{\beta}_{2SLS} = (\hat{X}'X)^{-1}\hat{X}'y.$$

   That is, $\hat{X}$ is used as an instrument for X;

2. Alternatively we can write as before:

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y,$$

   i.e., LS of $\hat{X}$ on y; or

3. We can write

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{y},$$

   where $\hat{y} = H_z y$. That is, residual regression of $\hat{y}$ on $\hat{X}$.

ASYMPTOTIC DISTRIBUTION OF 2SLS ESTIMATOR:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{D} N(0, \Lambda_{2SLS}),$$

where

$$\Lambda_{2SLS} = (\Pi_0'\Sigma_{zx})^{-1}\Pi_0'V_0\Pi_0(\Sigma_{zx}'\Pi_0)^{-1},$$

and V_0 is the asymptotic variance of $(1/\sqrt{n})Z'\epsilon$:

$$V_0 = \text{Va}(z_i\epsilon_i) = \lim_{n\to\infty} \text{Var}\left(\frac{1}{\sqrt{n}}Z'\epsilon\right)$$

and

$$\begin{aligned}
\Pi_0 &\equiv \operatorname*{plim}_{n\to\infty} (Z'Z)^{-1}Z'X \\
&= \left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}Z'Z\right)^{-1}\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}Z'X\right) \\
&= \Sigma_{zz}^{-1}\Sigma_{zx}.
\end{aligned}$$

So, the asymptotic covariance matrix of a 2SLS estimator is

$$\Lambda_{2SLS} = (\Sigma'_{zx}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}\Sigma'_{zx}\Sigma_{zz}^{-1}V_0\Sigma_{zz}^{-1}\Sigma_{zx}(\Sigma'_{zx}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}$$

SPECIAL CASE: Homoskedastic disturbance

Suppose that

$$\text{Var}(\epsilon \mid Z) = \sigma^2 I, \quad \Longrightarrow \quad V_0 = \sigma^2 \lim_{n\to\infty} \frac{1}{n} Z'Z = \sigma^2 \Sigma_{zz},$$

then Λ_{2SLS} simplifies to

$$\Lambda_{2SLS} = \sigma^2(\Sigma'_{zx}\Sigma_{zz}^{-1}\Sigma_{zx})^{-1}$$

and

$$\hat{\sigma}^2 \equiv \frac{1}{n}(y - X\hat{\beta})'(y - X\hat{\beta}) \xrightarrow{p} \sigma^2.$$

Note that

$$\frac{1}{n}\hat{X}'\hat{X} \xrightarrow{p} (\Sigma'_{zx}\Sigma_{zz}^{-1}\Sigma_{zx}),$$

where $\hat{X} = H_z X$. Thus,

$$\hat{\beta}_{2SLS} \overset{A}{\sim} N(\beta, \hat{\sigma}^2(\hat{X}'\hat{X})^{-1}).$$

REMARK: For the estimation of σ^2 we use X, not $\hat{X}$! We do use $\hat{X}$ only in $(\hat{X}'\hat{X})^{-1}$.

III. CHOICE OF AN OPTIMAL WEIGHT MATRIX

What is the optimal Π_0 so that the asymptotic variance of $\hat{\beta}_{2SLS}$ is minimized (in matrix sense)? That is, what is Π_0 so that

$$\Pi_0 \equiv \underset{\Pi}{\text{argmin}}\, (\Pi'\Sigma_{zx})^{-1}\Pi'V_0\Pi(\Sigma_{zx}\Pi)^{-1}.$$

Let Λ_0 be the asymptotic covariance with the optimal Π, then $\Lambda - \Lambda_0$ is positive semi definite matrix for any Π.

GLS REPRESENTATION:

$$y = X\beta + \epsilon.$$

$$\Longrightarrow \quad Z'y = Z'X\beta + Z'\epsilon.$$

$$\Longrightarrow \quad \frac{1}{\sqrt{n}}Z'y = \frac{1}{\sqrt{n}}Z'X\beta + \frac{1}{\sqrt{n}}Z'\epsilon,$$

or, the transformed model:

$$\tilde{y} = \tilde{X}\beta + \tilde{\epsilon},$$

where

$$\tilde{y} = \frac{1}{\sqrt{n}}Z'y, \quad \tilde{X} = \frac{1}{\sqrt{n}}Z'X, \quad \tilde{\epsilon} = \frac{1}{\sqrt{n}}Z'\epsilon.$$

Note that

$$\tilde{\epsilon} \xrightarrow{D} N(0, V_0),$$

by CLT.

Since $E[z_i\epsilon_i] = 0$, it is easy to show that

$$\text{Cov}(\tilde{X}, \tilde{\epsilon}) \xrightarrow{p} 0, \quad \text{as } n \to \infty.$$

Hence, the transformed model tends to GCR model, as $n \to \infty$.

So, define a GLS estimator by

$$\begin{aligned} \tilde{\beta}_{GLS} &= (\tilde{X}'V_0^{-1}\tilde{X})^{-1}\tilde{X}'V_0^{-1}\tilde{y} \\ &= \left(\frac{1}{n}X'ZV_0^{-1}Z'X\right)^{-1}\left(\frac{1}{n}X'ZV_0^{-1}\right)Z'y. \end{aligned}$$

This suggests that the optimal $\hat{\Pi}$ should be

$$\hat{\Pi}^* = \hat{V}_0^{-1}\left(\frac{1}{n}Z'X\right),$$

where

$$\hat{\Pi}^* \xrightarrow{p} \Pi^* \equiv V_0^{-1}\Sigma_{zx}$$

and

$$\hat{\beta}_{GMM} = \tilde{\beta}_{GLS} = (\hat{\Pi}^{*\prime}Z'X)^{-1}\hat{\Pi}^{*\prime}Z'y,$$

as before.

ASYMPTOTIC DISTRIBUTION OF GLS (GMM) ESTIMATOR:

If $\Pi^* \equiv V_0^{-1}\Sigma_{zx}$, then

$$\begin{aligned}\Lambda_{GLS} &= \left(\Pi^{*\prime}\Sigma_{zx}\right)^{-1}\Pi^{*\prime}V_0\Pi^{*\prime}(\Sigma_{zx}'\Pi^*)^{-1} \\ &= (\Sigma_{zx}'V_0^{-1}\Sigma_{zx})^{-1}.\end{aligned}$$

An immediate consequence of this result is that in general

$$\text{Va}(\hat{\beta}_{GMM}) \le \text{Va}(\hat{\beta}_{2SLS}),$$

in matrix sense.

REMARKS:

1. The 2SLS formula is very similar to GMM, just replace $(Z'Z)^{-1}$ in the 2SLS formula by V_0^{-1}.
2. In the homoskedastic case where $V_0 = \sigma^2\Sigma_{zz}$, we have
$$\text{Va}(\hat{\beta}_{GMM}) = \text{Va}(\hat{\beta}_{2SLS}).$$
3. The Generalized Method of Moments estimator is the best estimator which solves the sample analogue of the population moments
$$0 = E\left[\Pi z_i(y_i - x_i'\beta)\right].$$
4. In practice in order to obtain the GMM estimator, one needs to have a consistent estimator $\hat{V}$ of V_0. This can be done in two steps:

   Step 1: For initial estimator get the usual IV estimator using the weight matrix $\Pi = I \cdot \Sigma_{zx}$. Use $\hat{\epsilon}_i$ from this regression to form
$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2 z_i z_i'.$$

   Step 2: Use $\hat{V}$ from step 1 to get
$$\hat{\beta}_{FGMM} = \left(X'Z\hat{V}^{-1}Z'X\right)^{-1}X'Z\hat{V}^{-1}Z'y.$$